



Marketing Mix Model

Coursework for Applied Statistics and Business Forecasting:
Data Analysis & Regression Modelling on Cinnamon Tea Sales

CINNAMON TEA 11
10365619
BMAN 71791 – 2019/2020

Market Research: Background

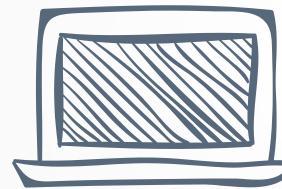
Consumer goods brand:
Cinnamon Tea



3 Years of Monthly Sales



5 Regional Markets:
North, South, East, West, Capital



3 Marketing Activities:
TV Ads, Online Banner, Direct Mail



Key Objectives:

- Analyse sales associations
- Evaluate marketing effectiveness
- Investigate possible sources of variation that drive sales

Dependent Variable : Sales
Independent Variables :

- | | |
|------------------|-----------|
| • Price | • Time |
| • TV Ads | • Product |
| • Online Ads | • Region |
| • Mail Promotion | • Month |

General Market Research of Cinnamon Tea

PRODUCT

Rich aroma tea with small unit size and simple packaging

PROMOTION

TV, newspapers, other media channels
Online ads, emails and social media marketing

MARKETING MIX

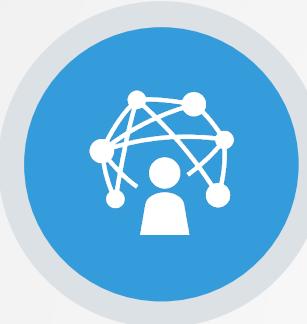
PRICE

Higher price for specialty tea (e.g. Cinnamon) than regular tea. Price range in UK: £2 - £5

PLACE

Distribution to supermarkets and retailers
Urban, Suburban and Rural areas

Method and Challenges



Steps:

1. Data Preparation
2. Splitting Training – Test Set
3. Creating Regression Model
4. Feature Selection: remove non-significant variables ($p\text{-value} > 0.05$)
5. Validation & Robustness Check
6. Analysis & Visualisation



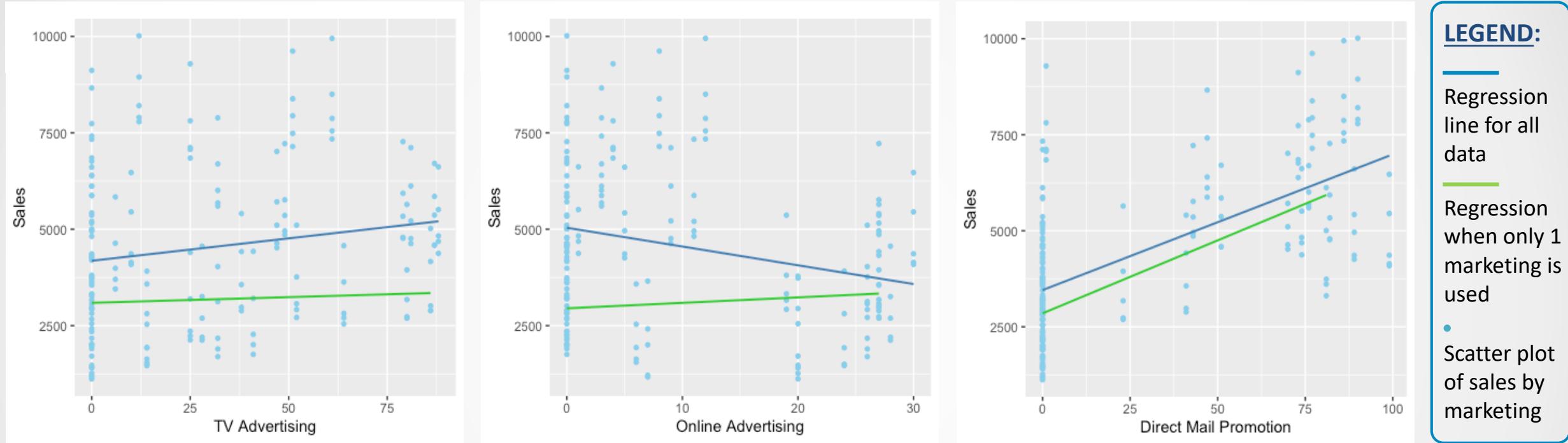
Challenges:

- Some information (e.g. exact year, exact location, cost of mail promotion) are not known.
- Variation of factors from variables such as customer's profile and demography of regions that are not available.
- Limited data (180 observations) to build model with many variables.

Method of Tests & Validation:

Tests: Significance, Residuals Diagnostics Plots, Multi-Collinearity, Auto-Correlation
Model Validation: Cross Validation

Data Visualisation: Simple Linear Regression of Marketing Activities



11.6 PACKS SOLD FOR EACH
TV ADVERTISING

64,064 IN TOTAL

2.9 PACKS WHEN ONLY TV AD IS USED

-48.6 PACKS SOLD FOR EACH
ONLINE ADVERTISING

-89,984 IN TOTAL

14.1 PACKS WHEN ONLY ONLINE AD IS USED

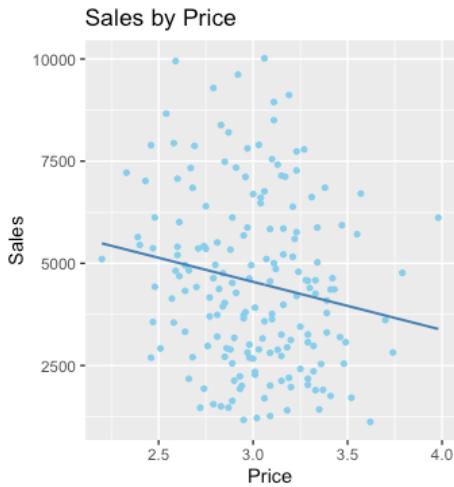
35.3 PACKS SOLD FOR EACH
DIRECT MAIL PROMOTION

194,910 IN TOTAL

38.0 PACKS WHEN ONLY DIRECT-MAIL IS USED

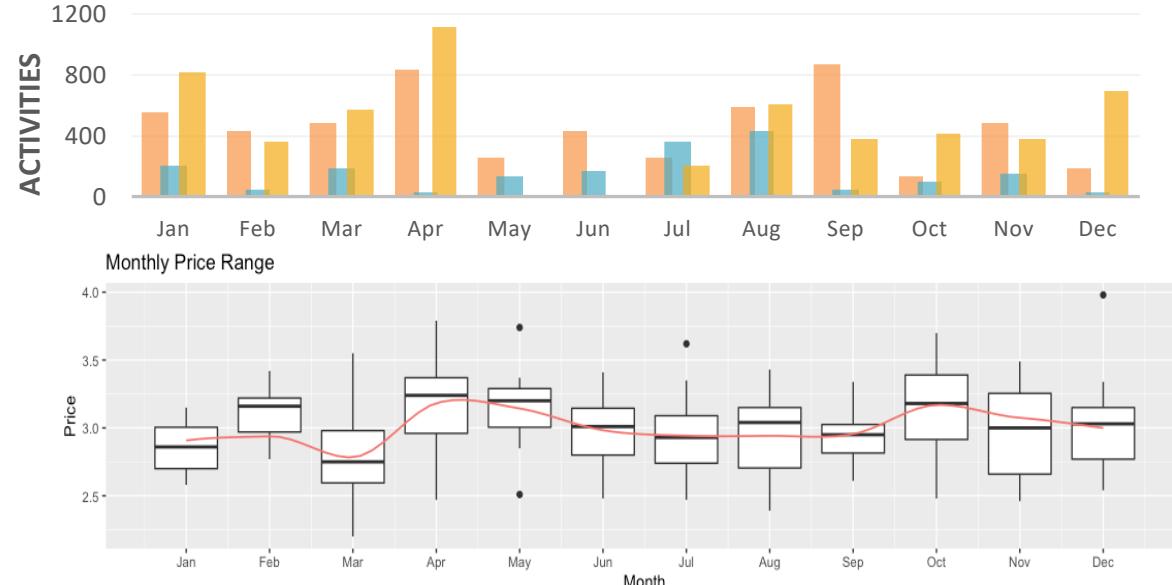
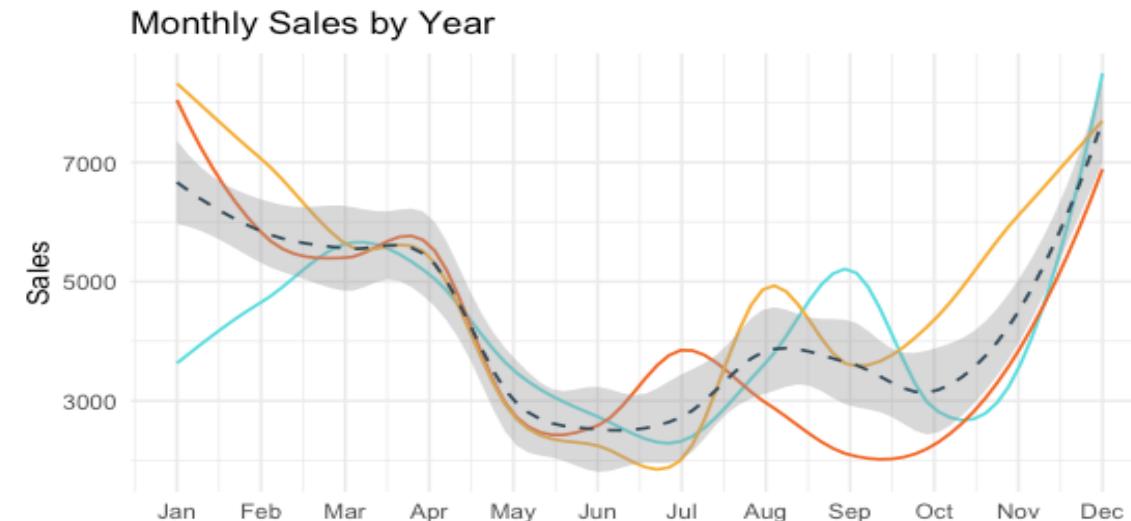
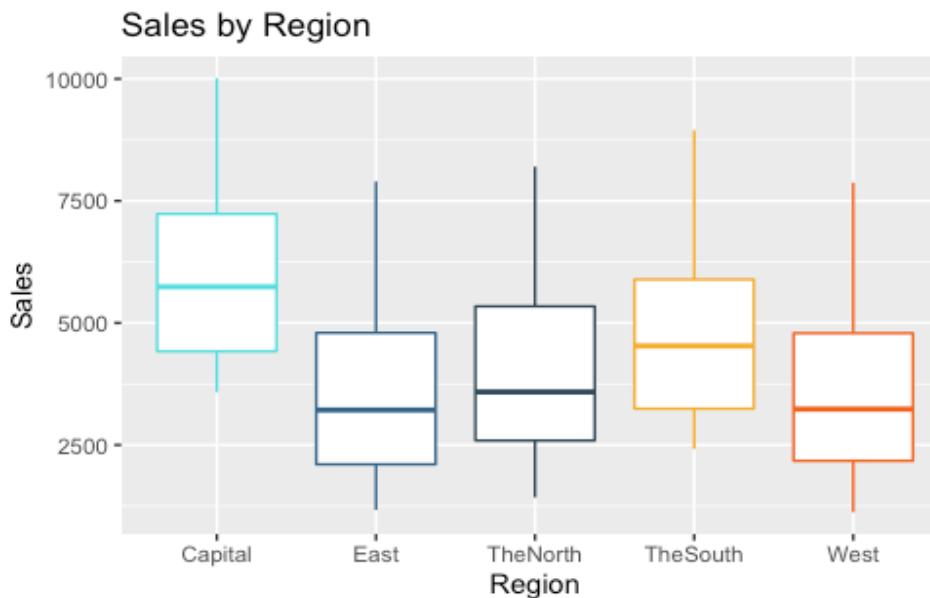
Regression with 3 activities as variables: **Sales associated = 4.83*(tv) – 25.62*(online) + 33.19*(mail) = 12.4**
!!! But they are not the best predictor of sales, more variables are needed !!!

Data Visualisation: Explore pattern/factors that drive sales



Highlight:

- Impact of a price increase is equal to sales decrease. Price is highest in Apr & Oct.
- Sales is highest in winter, specifically in December (month-wise) and in year 3.
- Direct-mail is the most used type of marketing but is never used in May & June.
- Capital has the highest sales.



Legend

- Average
- Year 1
- Year 2
- Year 3

Marketing per Month

- TV
- ONLINE
- MAIL

Average Price

Final Model: Multivariable Regression

Residuals:

	Min	1Q	Median	3Q	Max
	-872.81	-224.82	9.17	229.87	845.26

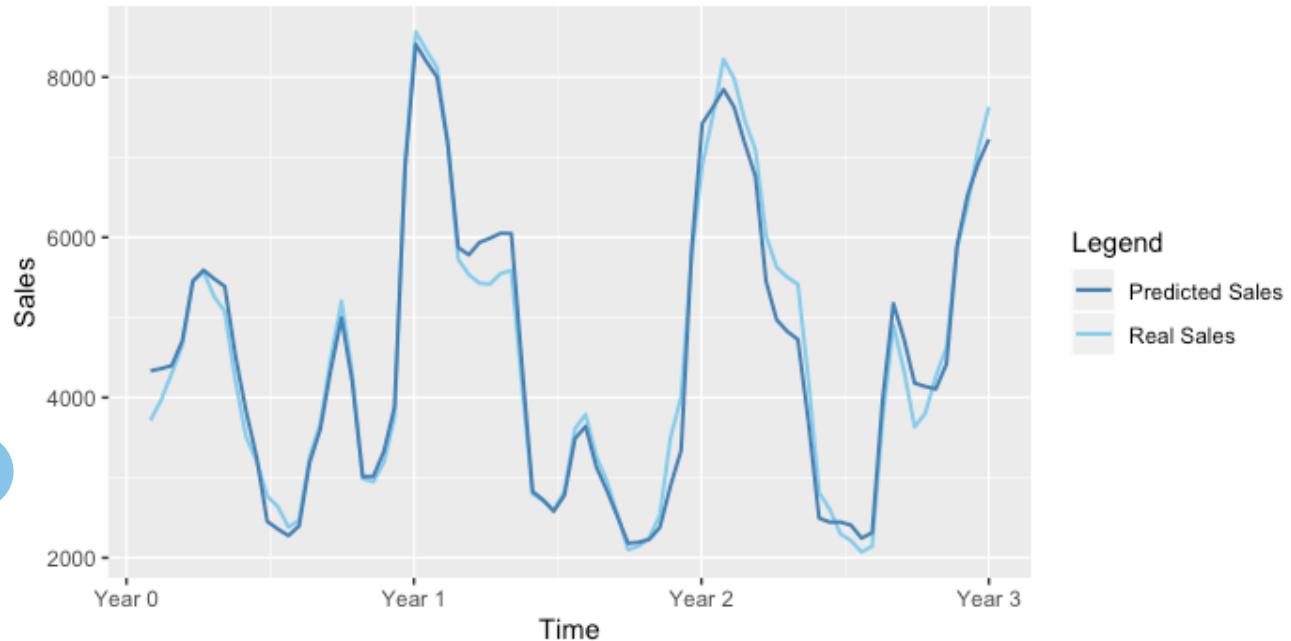
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2740.6622	355.0995	7.718	2.85e-12 ***
prom	27.8403	0.9886	28.160	< 2e-16 ***
Capital	2367.1926	81.5336	29.033	< 2e-16 ***
Dec	3845.7323	121.2572	31.715	< 2e-16 ***
prom_lastmonth	12.2767	0.9065	13.542	< 2e-16 ***
Jan	2037.7916	127.8261	15.942	< 2e-16 ***
Feb	2083.7761	123.3042	16.899	< 2e-16 ***
Mar	1667.4761	115.9226	14.384	< 2e-16 ***
South	1200.8077	84.7422	14.170	< 2e-16 ***
ad1	7.6923	1.1397	6.749	4.54e-10 ***
Nov	763.2777	125.5447	6.080	1.27e-08 ***
ad1_lastmonth	-5.1255	1.0996	-4.661	7.73e-06 ***
North	361.0742	90.4549	3.992	0.000109 ***
price	-400.1996	112.9754	-3.542	0.000553 ***
Apr	378.6523	148.5423	2.549	0.011971 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

2 Residual standard error: 367.9 on 129 degrees of freedom
Multiple R-squared: 0.9736, Adjusted R-squared: 0.9708
F-statistic: 340 on 14 and 129 DF, p-value: < 2.2e-16

Multivariable Regression of Cinnamon Tea Sales



1

1

2



Coefficient t-tests are significant: all 14 variables are not insignificant with 95% confidence

97.4% Goodness of fit & 97.1% Adjusted R²
The model is well-fitted with the observed value

Insights & Interpretation of Results



TV ads cost £2 mil: Effectiveness → **10%** ratio of revenue/cost**
Banners cost £500k: Effectiveness → **-55%*** ratio of revenue/cost**

TV ads is more effective than Online ads

*Online ads only shows positive impact when it's the only activity done

**Ratio = Total Sales generated from marketing x Mean of Price / Cost



Direct-Mail is the best marketing activity
with highest boost in sales + additional drive from direct-mail
promotion of the **previous month**



Sales of Cinnamon Tea has **seasonal cyclicalty**:
Peak sales & high correlation in **Dec & Jan (Winter)**
Sales is influenced by regions: **Capital, South & North**



When **price goes up, sales goes down**

To increase revenue, slightly increase price on high season,
decrease on summer and stabilise on the rest

From the final model,

Total packs of cinnamon tea sold
associated to marketing over 3 years:

195,924

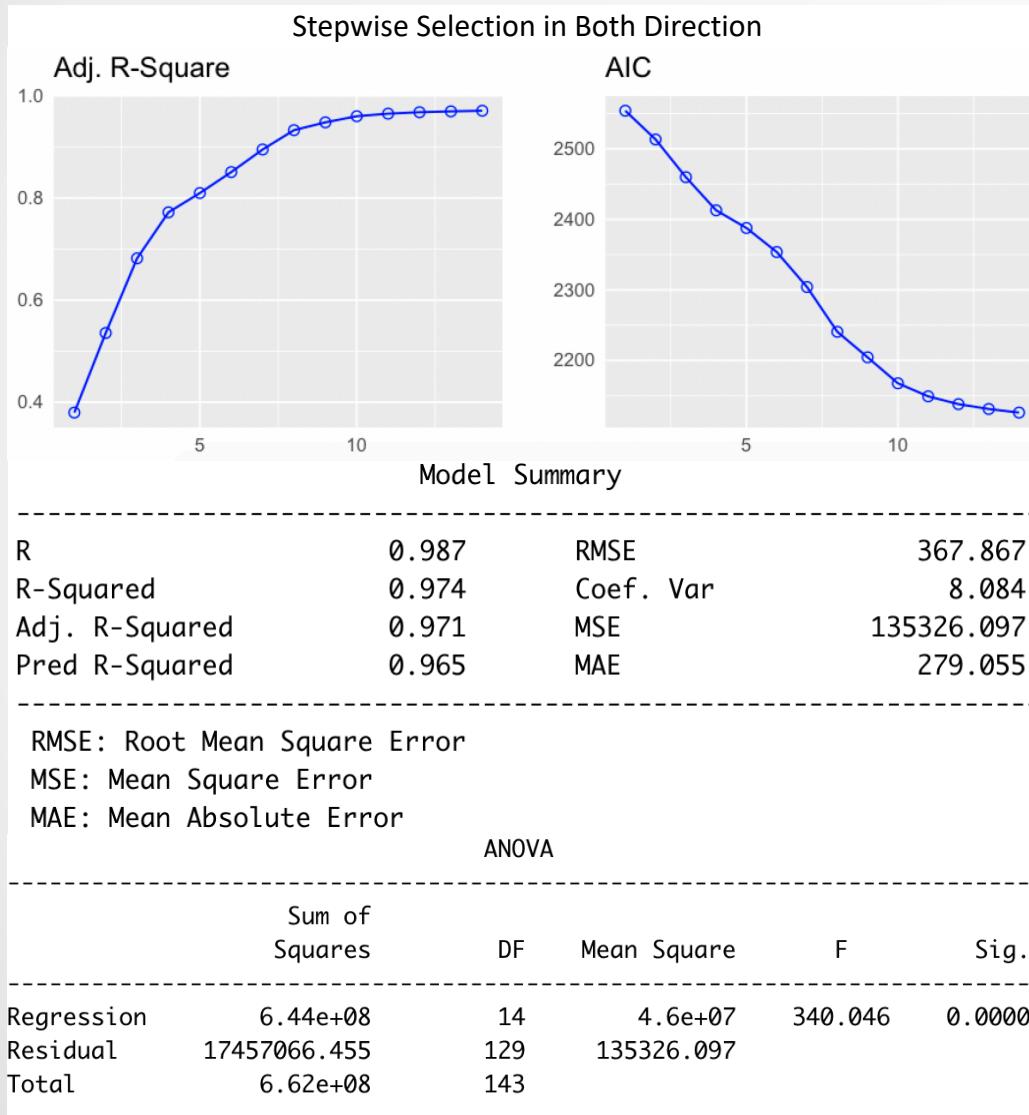
Each Mail Promotion: **27.84**

Each TV ads: **7.69**



Best to let go of online
banners & tv ads!
(they're not worth-it)
Boost mail promotion
**before and during
winter!**

Model Evaluation & Validation



Data Preparation, Splitting & Feature Selection:

1. Encoding categorical data into **binary variables**
2. Splitting Training-Test Set by **80-20**
3. Feature selection using Stepwise Selection Method: **14 predictors selected from 24 independent variables**



Validation Method:

K-Fold Cross Validation: Leave one out (LOOCV)

- **Adjusted R Squared = 97.1%**
- **Residual Standard Error (RSE) = 8.19%**
- Prediction Errors:
 - MSE = 135,326 ; RMSE = 367
 - MAE = 279



Overfitting & Robustness Check:

Bias vs Variance Trade-off: low bias & low variance = robust

- **Validation of Prediction Errors*:**
 - RMSPE Model (Predicted) = 348
 - RMSPE Cross Validation (cvpred) = 399
 - RMSPE Test Set Prediction = 412

The prediction error of the model, is relatively low and the difference of the value of error between that of Predicted, cvpred and the prediction on the test set are small

*Root Mean Square Prediction Error (RMSPE) $\sim \sqrt{\sum \text{Bias}^2/n}$

Assumption Tests & Limitations

Variance Inflation Factor (VIF)

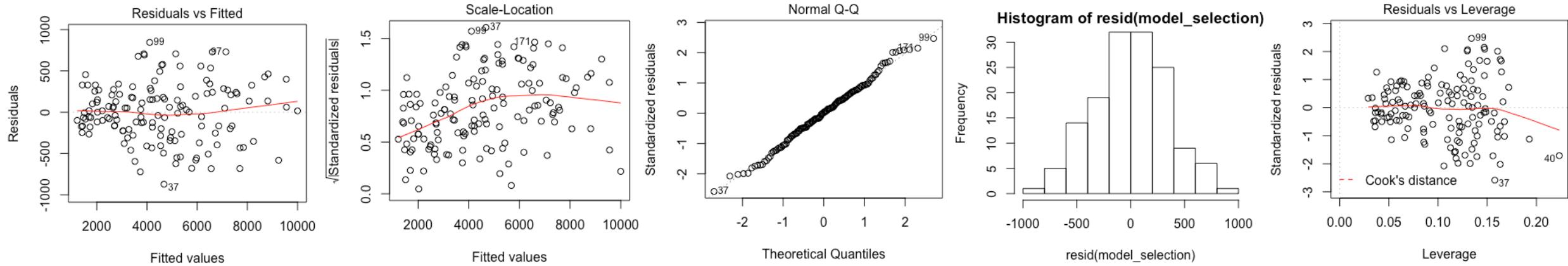
prom	Dec	Capital	prom_lastmonth	Jan	Mar	Feb
1.40	1.20	1.22	1.16	1.33	1.17	1.14
South	Nov	ad1	North	price	ad1_lastmonth	Apr
1.20	1.18	1.27	1.21	1.27	1.20	1.52

All variables have small VIF value, which means there's **no multi-collinearity** and all variables are independent

Durbin Watson

lag	Autocorrelation	D-W Statistic	p-value
1	0.176	1.63	0.018

D-W value < 2 which means there's small positive autocorrelation but still in **normal range between 1.5-2.5**
 p-value < 0.05 means we can reject null hypothesis



Residuals don't have non-linear relationships

Variance are equally spread along ranges of predictors, confirming homoscedasticity

Residuals are normally distributed

Variance of residuals has a relatively normal distribution

There's no outlier that is influential to the model

Ordinary Least Squares (OLS) assumptions are overall met. Several remaining notes: small positive autocorrelation which shows that there might be trends over time and some other factors might lead to slightly heteroscedastic residuals for smaller fitted values, but still acceptable. Forecasting sales based on past data **relies on the assumption that the market pattern will remain constant** over time. However, external effects, such as launching of competitor's products, events, economic crises etc, may also affect sales.

References

Alloy Client Solutions. (2019). *The Role Of Seasonality In Demand Forecasting Success.*

<https://medium.com/alloytech/the-role-of-seasonality-in-demand-forecasting-success-171166a7039>

Anderson, D., Sweeney, D., Williams, T. (2011). *Statistics for Business and Economics*. South-Western, Cengage Learning.

Kim, Bommae. (2015). *Understanding Diagnostic Plots for Linear Regression Analysis*. University of Virginia Library Research Data Services + Sciences. <https://data.library.virginia.edu/diagnostic-plots/>

MBASkool. *Lipton Marketing Mix (4Ps) Strategy*. <https://www.mbaSkool.com/marketing-mix/products/17520-lipton.html>